

Semantic similarity and analysis of the word frequency dynamics

Bochkarev V., Maslennikova Y., Svetovidov A.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© Published under licence by IOP Publishing Ltd. In this study a similarity in changes of frequencies dynamics for semantically related words was analyzed using word statistics extracted from more than 4.5 million books written over a period of 205 years. The approach is based on the correlation analysis of 1-grams frequency dynamics. We analyzed the frequencies correlation of synonym pairs, their corresponding antonymous groups and random words pairs. Also, we compared several metrics to find the most effective for assessing the degree of similarity in the dynamics of use of different words. Comparing differences between logarithmic rank variations in pairs of synonyms and random word pairs, significant differences are found, though they are smaller than it could be expected.

<http://dx.doi.org/10.1088/1742-6596/936/1/012067>

References

- [1] Leeuwenberg A, Vela M, Dehdari J and Genabith J A 2016 Minimally Supervised Approach for Synonym Extraction with Word Embeddings *The Prague Bulletin of Mathematical Linguistics* 111-142
- [2] Liu D and Zhong S 2016 L2 vs. L1 use of synonymy: An empirical study of synonym use/acquisition *Applied Linguistics* 37 239-261
- [3] Steffen Roth et al and Mesquita A. 2016 *Research Paradigms and Contemporary Perspectives on Human-Technology Interaction* (Hershey: IGI-Global) *The Fashionable Functions Reloaded: An Updated Google Ngram View of Trends in Functional Differentiation (1800-2000)*
- [4] Montemurro M A and Zanette D H 2016 Coherent oscillations in word-use data from 1700 to 2008 *Palgrave communications*
- [5] Michel J, Shen Y, Aiden A et al 2011 Quantitative Analysis of Culture Using Millions of Digitized Books *Science* 331 176-182
- [6] Lin Y, Michel J-B, Aiden EL, Orwant J, Brockman W and Petrov S 2012 Syntactic annotations for the google books ngram corpus, in 'Proceedings of the ACL
- [7] Fellbaum C. 1998 *WordNet: An Electronic Lexical Database* (Cambridge, MA: MIT Press)
- [8] Bochkarev V, Solovyev V and Wichmann S 2014 Universals versus historical contingencies in lexical evolution *J. R. Soc. Interface* 11 0841 2014